Composite likelihood for space-time data\textsuperscript{1}

Carlo Gaetan

Paris, May 11th, 2015

\textsuperscript{1}I thank my colleague Cristiano Varin for sharing ideas, slides and office with me.
Bruce Lindsay

March 7, 1947 - May 5, 2015

Contemporary Mathematics 80, 221-239
General framework

- In several cases models have complex interdependencies and/or we deal with large dataset.
- Joint distribution of the data may be difficult to evaluate, or even to specify!
- Typical difficulties arise from the need to invert large matrices when we deal with large dataset and/or from approximation of intractable integrals.
- Ancient Roman’s principle: dividi et impera
  \( \text{`diviser pour regner', Louis XI } \)
- Idea: if computing likelihoods for certain subsets of the data is possible, then one may construct a pseudolikelihood by combining such likelihood objects and use this as a surrogate for the ordinary likelihood.
- Review paper: Varin et al. (2011)
Example: spatio-temporal Gaussian data

- $s \in \mathbb{R}^d$ is a spatial location, $t \in \mathbb{R}$ is a time point
- Observations $z = (z(s_1, t_1), \ldots, z(s_n, t_n))'$ from a Gaussian Random Field (GRF) $\{Z(s, t)\}$
- Weakly stationarity

\[
\text{cov}(Z(s, t), Z(s', t')) = C(h, u; \theta)
\]

($h = s - s'$, spatial lag, $u = t - t'$ temporal lag).
- Likelihood computation requires inversion of $n \times n$ covariance
  - $O(n^3)$ operations, $O(n^2)$ memory
  - If $n$ large, this may be unfeasible!
Two possible strategies:

1. simplify the model:
   - approximate with a Gaussian Markov Random Field (Lindgren et al., 2011) requiring roughly $O(n \log n)$ operations
   - approximate with low rank methods (Cressie and Johannesson, 2008) $O(n)$

2. keep your model but simplify the fitting method: e.g. Curriero and Lele (1999) estimate $\theta$ from the pseudolikelihood

$$PL(\theta; z) = \prod_{i>j} f(z(s_i, t_i) - z(s_j, t_j); \theta) w(s_i, t_j, s_j, t_j)$$
Example: spatio-temporal non-gaussian data I

- Generalized linear geostatistical models (latent Gaussian models)

\[ E(Z(s, t)|U(s, t)) = g^{-1}(x(s, t)'\beta + U(s, t)) \]

where \( \{U(s, t)\} \) is a latent GRF.

- In general, the likelihood involves an n-dimensional integral

\[ L(\beta, \psi, \theta; z) = \int_{\mathbb{R}^n} f(z|u; \beta, \psi)f(u; \theta)du \]

- Monte Carlo methods: MCEM, MCMC, etc. may be time-consuming even for moderate \( n \)
  - INLA method (Rue et al., 2009) ⇒ see Thomas’s talk

- Simpler: pseudolikelihood constructed from bivariate margins

\[ PL(\theta; z) = \prod_{i>j} f(z(s_i, t_i), z(s_j, t_j); \theta)^{w(s_i, t_i, s_j, t_j)} \]
Example: spatio-temporal extreme values

- Observations $z = (z(s_1, t_1), \ldots, z(s_n, t_n))'$ are maxima recorded values.
- Classical extreme value theory says that the marginal distribution can be modeled by the GEV distribution

$$\Pr(Z(s, t) \leq z) = \exp \left[ -\left\{ 1 + \xi (z - \mu) / \sigma \right\}^{-1/\xi} \right]$$

with $\mu \in \mathbb{R}$, $\xi \in \mathbb{R}$, $\sigma > 0$ and $\{1 + \xi (z - \mu) / \sigma\} > 0$.
- Unit Fréchet margins (i.e. $\mu = \xi = \sigma = 1$)

$$\Pr(Z(s, t) \leq z) = \exp \left[ -1/z \right]$$
Example: spatio-temporal extreme values II

- A **max-stable process** (de Haan, 1984), is the extension for maxima of independent replications of (space-time) random fields.
- Suppose unit Fréchet margins. Finite dimensional distribution

\[
Pr(Z \leq z) = Pr(Z(s_1, t_1) \leq z_1, \ldots, Z(s_n, t_n) \leq z_n) = \exp(-V(z))
\]

\(V(z)\) is a positive function such that \(V(a^{-1}z) = aV(z)\) for any \(a > 0\) and \(z > 0\)

- Assume a parametric model

\[
Pr(Z \leq z) = \exp(-V(z; \theta))
\]

the likelihood corresponds to the derivative with respect to all components of \(z\).
- The number of terms is the \(B_n\) Bell number, around \(8.3 \times 10^{10}\) for \(n = 17\)!
Consider

1. a statistical model \( \{ f(z; \theta), z \in \mathbb{R}^n, \theta \in \Theta \subseteq \mathbb{R}^p \} \);
2. a set of measurable events \( \{ A_k; k = 1, \ldots, K \} \);
3. the associated likelihoods \( \mathcal{L}_k(\theta; z) = f(z \in A_k; \theta) \).

Then, a composite likelihood (CL) is the weighted product of the likelihoods corresponding to each single event,

\[
\text{CL}(\theta; z) = \prod_{k=1}^{K} \mathcal{L}_k(\theta; z)^{w_k},
\]

where \( \{ w_k; k = 1, \ldots, K \} \) are positive weights.
Composite conditional likelihood

- Notation $z_i = z(s_i, t_i)$, $z_D = \{z_j, j \in D\}$, $D \subset \{1, \ldots, n\}$ and $D$ is a set of $D$.

$$CCL(\theta; z) = \prod_{D \in D} f(z_D|z_{D^c}; \theta),$$

- Besag's pseudolikelihood (Besag, 1974)

$$CCL(\theta; z) = \prod_{i=1}^{n} f(z(s_i)|z_{\partial_i}; \theta),$$

$\partial_i$ neighborhood of $s_i$


- Limited number of space-time applications
  - Mixed states spatio-temporal auto-models on regular lattice (Hardouin and Crivelli, 2011)
Composite marginal likelihood

- More commonly used

\[ CML(\theta; z) = \prod_{D \in \mathcal{D}} f(z_D; \theta) \]

- Independence likelihood (Smith, 1990; Chandler and Bate, 2007) where the sets \( D \) are disjoint
  - \( \theta = (\theta_S, \theta_T)' \) and inference about \( \theta_S \)

\[ CML(\theta; z) = \prod_{t=1}^{T} f(z(t); \theta_S) \]

with \( z(t) = (z(s_1, t), \ldots, z(s_k, t)) \)

- pairwise likelihood: \( \prod_{i>j} f(z_i, z_j; \theta) \)

- tripletwise likelihood, \ldots, ”blockwise” likelihood

- pairwise differences (for GRF): \( \prod_{i>j} f(z_i - z_j; \theta) \)
Hybrid methods

- **hybrid pairwise likelihood**: Kuk (2007)
  - optimal estimating equations for marginal parameters
  - pairwise likelihoods for estimating dependence parameters
  - no examples for space-time examples yet!

- **joint composite estimating functions**: Bai et al. (2012)
  - space-time example, see later.
Composite likelihood quantities

- **Composite log-likelihood**

\[
cl(\theta; \mathbf{z}) = \log CL(\theta; \mathbf{z}) = \sum_{k} \log \ell_k(\theta; \mathbf{z}) w_k
\]

- **Composite score**

\[
u(\theta; \mathbf{z}) = \sum_{k=1}^{K} \log \nabla \ell_k(\theta; \mathbf{z}) w_k = \sum_{k=1}^{K} u_k(\theta; \mathbf{z}) w_k, \Rightarrow E[u(\theta; \mathbf{Z})] = 0
\]

- **Maximum composite estimator** \( \hat{\theta}_{CL} \)

\[
cl(\hat{\theta}_{CL}; \mathbf{z}) \geq cl(\theta; \mathbf{z}), \quad \forall \theta \in \Theta
\]
Composite likelihood quantities II

- Godambe (or sandwich) information

\[ G(\theta) = H(\theta)J(\theta)^{-1}H(\theta) \]

- Sensitivity matrix: \( H(\theta) = E[-\nabla u(\theta; Z)] \),

\[ \hat{H} = -\nabla u(\hat{\theta}_{CL}; z) \]

- Variability matrix: \( J(\theta) = \text{Var}[u(\theta; Z)] \)

  How to estimate?

- Misspecification \( H(\theta) \neq J(\theta) \)
Asymptotics I

- Two types of asymptotic frameworks
  - fixed domain
    - fill a space-time region with a sequence of observed locations
    - no results
  - increasing domain:
    - expand a space-time region for recover more observed locations
    - regular lattice: \( R_0 \subset (-\frac{1}{2}, \frac{1}{2}]^{d+1} \),
      \[ R_n = \{(s_1, t_1), \ldots, (s_n, t_n)\} = (\lambda_n R_0) \cap \mathbb{Z}^{d+1} \]
Asymptotics II

- Idea: log-composite likelihood is an additive contrast function (Hardouin, 1992; Guyon, 1995)

\[ U_n(\theta) = \frac{1}{K_n} \sum_{k=1}^{K_n} \log f(z \in A_k; \theta)w_k \]

- mixing conditions on \( \{Z(s, t)\} \)
- \( \hat{\theta}_{CL} \) is asymptotically Gaussian

\[ G(\theta)^{1/2}(\hat{\theta}_{CL} - \theta^*) \xrightarrow{d} \mathcal{N}(0, I) \]

- \( \theta^* \) is the minimizer of the composite Kullback-Leibler divergence

\[ CKL(\theta; g, f) = \sum_k E_g \{\log g(Z \in A_k) - \log f(Z \in A_k; \theta)\} w_k \]
Asymptotics III

• consistency if all composite likelihood ”blocks” correctly specified

\[ \exists \theta \text{ such that } f(z \in A_k; \theta) = g(z \in A_k) \quad \forall k \in K \]

• Space-time examples:
  - **GRFs**: Bevilacqua et al. (2012); Bai et al. (2012)
  - **Max-stable processes**: Davis et al. (2013)
Hypothesis testing

Null hypothesis $H_0 : \psi = \psi_0$ where $\theta = (\psi, \tau)$ and $\psi \in \mathbb{R}^q$

- Wald-type statistics (no invariance under reparameterization)
- Score-type statistics (numerical instability)
- Composite likelihood ratio statistics

$$\mathcal{W} = 2\{c\ell(\hat{\theta}_{CL}; z) - c\ell(\psi_0, \hat{\tau}_{CL}(\psi_0); z)\}$$

▶ Non standard distribution

$$\mathcal{W} \xrightarrow{d} \sum_{j=1}^{q} \lambda_j \chi^2_{1,j}$$

$\lambda_1 \geq \ldots \geq \lambda_q$ eigenvalues of $(H_{\psi\psi})^{-1} G_{\psi\psi}$ (Kent, 1982; Guyon, 1995)

▶ How to calibrate $\mathcal{W}$? Spatial example in Cattelan and Sartori (2014)
Model selection

- Model selection based on Akaike-type criterion (Takeuchi, 1976)
- Composite likelihood information criterion (Varin and Vidoni, 2005)
  \[ \text{CLIC} = -2cl(\hat{\theta}_{CL}; z) + 2 \dim(\theta) \]
- Composite Bayesian information criterion (Gao and Song, 2010)
  \[ \text{CBIC} = -2cl(\hat{\theta}_{CL}; z) + \log(n) \dim(\theta) \]
- Effective number of parameters
  \[ \dim(\theta) = \text{tr}(H(\theta)J^{-1}(\theta)) \]
- CLIC has a tendency to select over-complicated models.
Model selection: example I

- 100 independent simulations from a zero mean space-time Gaussian process with covariance function $C(h, u; \theta)$
- two space-time grids: $S \times T$, where $S = \{-2, \ldots, 3\}^2$ and $T = \{1, \ldots, T\}$ with $T = 100, 200$,
- $Z(s_i, t_i) - Z(s_j, t_j) \sim \mathcal{N}(0, 2\gamma(s_i - s_j, t_i - t_j))$ with $\gamma(h, u) = C(0, 0; \theta) - C(h, u; \theta)$
- pairwise difference likelihood

\[
PL(\theta; z) = \prod_{i>j} f(z(s_i, t_i) - z(s_j, t_j); \theta)
\]
Model selection: example II

model A  Double exponential model (separable model)

\[ C(h, u; \theta) = \sigma^2 \exp \left( -3 \frac{\|h\|}{a} - 3 \frac{|u|}{b} \right) \]

model B  Gneiting model

\[ C(h, u; \theta) = \frac{\sigma^2}{\left( \frac{20 |u|^{2\alpha}}{b} + 1 \right)} \exp \left\{ - \frac{3\|h\|}{a \left( \frac{20 |u|^{2\alpha}}{b} + 1 \right)^{\beta/2}} \right\} , \]

with \( \beta = 0 \) (separable model)

model C  Gneiting model with \( \beta \neq 0 \);
Model selection: example III

<table>
<thead>
<tr>
<th>Model</th>
<th>(A)</th>
<th>(B)</th>
<th>(C)</th>
<th>(A)</th>
<th>(B)</th>
<th>(C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td>84</td>
<td>10</td>
<td>6</td>
<td>96</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>(B)</td>
<td>8</td>
<td>70</td>
<td>22</td>
<td>5</td>
<td>82</td>
<td>13</td>
</tr>
<tr>
<td>(C)</td>
<td>9</td>
<td>19</td>
<td>72</td>
<td>3</td>
<td>12</td>
<td>85</td>
</tr>
</tbody>
</table>

- at least 70% of the models have been correctly identified.
Large data set: around 150000 observations
variability matrix $J(\theta) = \text{Var}[u(\theta; Z)]$ key ingredient for
- evaluating uncertainty of composite likelihood estimates
- test statistics
- information criteria for model selection

the simpler case is when there are $T$ independent replicates (in time, for instance) $z(t) = (z(s_1, t), \ldots, z(s_k, t))$

$$
\hat{J} = \frac{1}{T} \sum_{i=1}^{T} u(\hat{\theta}_{MC}; z(t))u(\hat{\theta}_{MC}; z(t))^T
$$

space-time models for extreme values (Huser and Davison, 2014)
more stable to estimate directly the variance matrix of $\hat{\theta}_{CL}$ by bootstrap but computational demanding (Bags of little bootstrap ?, Kleiner et al. (2014))

● much harder when there is no replication in space or time

● analytic expressions are possible only for few special cases (spatial GRF, Bevilacqua and Gaetan (2014)) but too computational demanding ($O(n^4)$)

● resampling methods based on idea of some form of ”internal replication” (block-bootstrap, window subsampling)
  ▶ difficult to specify tuning parameters (block dimension)
  ▶ arduous with irregularly spaced temporal and spatial observations
Composite likelihood design I

- How to choose among many possible composite likelihoods?
- for example, which is preferable between
  - the pairwise marginal likelihood
    \[ c_{\ell_m}(\theta; z) = \sum_{i > j} w_{ij} \log f(z_i, z_j; \theta) \]
  - the conditional likelihood
    \[ c_{\ell_c}(\theta; z) = \sum_{i \neq j} w_{ij} \log f(z_i | z_j; \theta) \]
  - the pairwise difference likelihood (for GRF)
    \[ c_{\ell_d}(\theta; z) = \sum_{i > j} w_{ij} \log f(z_i - z_j; \theta) \]

- Same computational burden: \( O(n^2) \)
- In large settings: should all possible pairs be included?
Weighted composite likelihoods

- When efficiency is low, unequal weighting of likelihood components can be used to obtain some improvement.
- Optimal weights computationally impractical.
- It sensible to assume $w_{ij} = w_{ji}$ for $c_{\ell m}$ and $c_{\ell d}$.
- If $w_{ij} = w_{ji}$ and $w_{ii} = 1$

\[
c_{\ell c}(\theta; z) = 2c_{\ell m}(\theta; z) - (n - 1) \sum_{i=1}^{n} \log f(z_i; \theta)
\]

- When the marginal parameters are known, marginal and the conditional pairwise likelihood have the same efficiency.

- in space-time setting underweight observations that are far apart in time and/or space.
'Simple' weighted composite likelihoods I

- weights based on the distances.

\[ w_{ij}(d) = \begin{cases} 
1 & \|s_i - s_j\| \leq d_s, |t_i - t_j| \leq d_t, \quad d = (d_s, d_t)' \\
0 & \text{otherwise}
\]
'Simple' weighted composite likelihoods II

- Spatial setting results (Bevilacqua and Gaetan, 2014)

---

![Graph showing semi-variogram models]

### Table 3: Bias and root mean square errors (rmse) of the estimates when $\mu = 0$, $\phi = 0$ and $\sigma^2 = 1$.

<table>
<thead>
<tr>
<th></th>
<th>Exponential model</th>
<th>Cauchy model</th>
<th>Wave model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$</td>
<td>$\phi$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>$ML$</td>
<td>bias</td>
<td>0.0037</td>
<td>-0.0015</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.0843</td>
<td>0.0114</td>
</tr>
<tr>
<td>$TAP$</td>
<td>bias</td>
<td>0.0039</td>
<td>-0.0013</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.0851</td>
<td>0.0129</td>
</tr>
<tr>
<td>$pl_C$</td>
<td>bias</td>
<td>0.0041</td>
<td>-0.0028</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.0900</td>
<td>0.0133</td>
</tr>
<tr>
<td>$pl_M$</td>
<td>bias</td>
<td>0.0041</td>
<td>-0.0028</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.0901</td>
<td>0.0133</td>
</tr>
<tr>
<td>$pl_D$</td>
<td>bias</td>
<td>0.0011</td>
<td>0.0088</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.0201</td>
<td>0.1142</td>
</tr>
</tbody>
</table>
'Simple’ weighted composite likelihoods III

- Practical implementation (Bevilacqua et al., 2012)
  - Get a consistent estimate for $\theta$, $\tilde{\theta}$
  - Estimate $G(\theta, d)$ by $\hat{G}(\tilde{\theta}; d)$
  - We choose the ‘lag’ $d$ minimizing $[\hat{G}(\tilde{\theta}; d)]^{-1}$ in the partial order of nonnegative definite matrices or equivalently

$$d^* = \arg\min_{d \in \mathcal{D}} \text{tr}[\hat{G}(\tilde{\theta}; d)]^{-1}$$

where $\mathcal{D}$ is a set of space-time lags.

- Maximize

$$c_{lm}(\theta; z) = \sum_{i > j} w_{ij}(d^*) \log f(z_i, z_j; \theta)$$
Joint composite estimating functions

- First partition $R = \{(s_1, t_1), \ldots, (s_n, t_n)\}$ into three subsets
  - $R_S$ with pairs $(s_i, t_i), (s_j, t_j)$, $i \neq j$ differing only in locations i.e. $s_i \neq s_j$, and $t_i = t_j$
  - $R_T$ with pairs differing only in time i.e. $s_i = s_j$, and $t_i \neq t_j$
  - $R_C$ with pairs (cross-pairs) differing in time and in space i.e. $s_i \neq s_j$, and $t_i \neq t_j$

\[ \Psi_{CL.\theta/} = \sum_{k \in D_n} f_k \{d.k/; \theta\}, \]

where $d.k/\,$ are implicitly treated as being independent.

Alternatively, one may stack the individual composite score function terms into a column vector $\nu$. $\nu/\theta = \{f_k \{d.k/; \theta\}\}_{k \in D_n}$, from which the estimating function is given by $E\{\dot{\nu} \cdot \theta/\} T \text{cov} \{\nu \cdot \theta/\}^{-1} \nu \cdot \theta/\} = 0$.

As pointed out by Kuk (2007), this version of composite estimating equations effectively accounts for the correlations between the differences. However, the calculation of $\text{cov} \{\nu \cdot \theta/\}$ and its inverse is computationally prohibitive when the number of pairs (or differences) is large.

To improve on the existing CL methods and to incorporate correlations between the pairs in the estimation, we propose a new approach, i.e. we construct three sets of estimating functions by using the spatiotemporal characteristics of the data. Specifically, we first partition $D_n$ into three subsets, namely $DS_n$, $DT_n$, with pairs differing only in locations, $DT_n$, with pairs differing only in time and $DC_n$, with pairs (cross-pairs) differing in time and in space.

Fig. 1 displays such a partition with the three types of pairs, (a) for a spatial pair, (b) for a temporal pair and (c) for a spatiotemporal cross-pair, in a typical spatiotemporal setting with four locations observed at two time points.

Summing over all pairwise differences of spatial pairs across all time points, we obtain the following spatial composite estimating function (CEF):

\[ \Psi_{S,n.\theta/} = \frac{1}{|DS_n|} \sum_{i \in DS_n} f_i \{d.i/; \theta\}, \]

where $|A|$ denotes the number of elements in $A$.

In a similar fashion, we construct the temporal CEF:

\[ \Psi_{T,n.\theta/} = \frac{1}{|DT_n|} \sum_{i \in DT_n} f_i \{d.i/; \theta\}, \]

From which the estimating function is given by $E\{\dot{\nu} \cdot \theta/\} T \text{cov} \{\nu \cdot \theta/\}^{-1} \nu \cdot \theta/\} = 0$.

Fig. 1.

Configurations of spatiotemporal pairs: the upper plane represents four locations observed at time 1, and the lower plane represents the same four locations observed at time 2; (a) is the spatial pair, (b) the temporal pair and (c) the spatiotemporal cross-pair.
Joint composite estimating functions II

- Define $c_\ell_A(\theta; \mathbf{z}) = \sum_{(i,j) \in A} w_{ij} \log f(z_i, z_j; \theta)$ we have

$$c_\ell_m(\theta; \mathbf{z}) = c_\ell_S(\theta; \mathbf{z}) + c_\ell_T(\theta; \mathbf{z}) + c_\ell_C(\theta; \mathbf{z})$$

complete = space + time + space-time

- Composite likelihood estimate $\hat{\theta}_{CL}$ solve the estimating equation

$$u(\theta; \mathbf{z}) = 0$$

$$u_S(\theta; \mathbf{z}) + u_T(\theta; \mathbf{z}) + u_C(\theta; \mathbf{z}) = 0$$
Joint composite estimating functions III

- Instead stack the subset composite score functions into the vector \( g(\theta, z) = (u_S(\theta; z), u_S(\theta; z), u_C(\theta; z)) \)

- Weighted quadratic objective function (looks like GMM (Hansen, 1982))

\[
Q(\theta; z) = g(\theta, z)^\top W(\theta) g(\theta, z)
\]

where \( W(\theta) = D \text{Var}[g(\theta, Z)] D \) and
\[
D = \text{diag}(\sqrt{|R_S|}, \sqrt{|R_T|}, \sqrt{|R_C|}) \otimes I \) (adjustment for different size)

- The joint composite estimating function (JCEF) estimator (Bai et al., 2012)

\[
\hat{\theta}_{JCEF} = \arg\max_\theta Q(\theta; z)
\]
Joint composite estimating functions IV

- More efficient than 'simple' weighted composite likelihood for a fixed $d$
- More computational demanding
  - Require at each time an evaluation of $\text{Var}[g(\theta, Z)]$
  - $\text{Var}[g(\theta, Z)]$ can be derived analytically for GRF, given the large number of possible pairs, computing it on the basis of analytic formulae is not practically feasible.
  - estimation of this matrix is typically achieved by subsampling techniques
- Model selection and hypothesis testing ?
Big data: block composite likelihoods

- Estimation and prediction for spatial data (Eidsvik et al., 2014)
- A framework that allows parallel computing
- Similar approaches: Vecchia (1988); Stein et al. (2004); Caragea and Smith (2006).
- Partition a region $D$ into $M$ blocks $D_1, \ldots, D_M$, denote $z_{D_k} = \{z_i, i \in D_k\}$
- Provide that the number of locations in $D_k \cup D_l$ is no so large for evaluating $f(z_{D_k}, z_{D_l}; \theta)$...

$$c\ell_B(\theta; z) = \sum_{l>k} \log f(z_{D_k}, z_{D_l}; \theta)$$

(log) block composite likelihood (pairwise composite block-likelihood)
Big data: block composite likelihoods II

- for $M = 1$ or $M = 2$ full likelihood, $M = n$ pairwise likelihood
- $M$: trade-off between computational and statistical efficiency.
- $N_k$ the neighbors of block $k$, $N_k \rightarrow = \{l > k\} \cap \{l \in N_k\}$

\[
c\ell_B(\theta; z) = \sum_{k=1}^{M-1} \sum_{l \in N_k} \log f(z_{D_k}, z_{D_l}; \theta)
\]

Source: Eidsvik et al. (2014)
Recommendation for creating blocks

- check the spatial dependence and possible anisotropies with empirical variogram

Computational efficiency

- $n_k = c$ (fixed number of observations in a block)
- $O(c^3 M | N_k \rightarrow |)$ i.e. $O(n)$

Concurrent approach:

- Fixed-rank kriging (Cressie and Johannesson, 2008): $O(n)$
- Gaussian Markov random fields (Lindgren et al., 2011): $O(n^{3/2})$ for two-dimensional spatial data and $O(n^2)$ for three spatial dimensions
Software development (in R)

- is R a convenient framework for development of composite likelihood software?
- coding in R: (block) pairwise (marginal or conditional) likelihood needs parallelizable bivariate functions in way to avoid slow loops
- Graphical processing unit
- more efficient (?) coding composite likelihoods and derived quantities in low-level languages such as Fortran or C and then call into R
- efficient numerical algorithms for low-dimension integration (for non Gaussian data)
- what is already available in R: CompRandFld package for GRF
Challenges as final remarks

• "Design" issues: which terms should be included and how they have to be combined?
  ▶ Preliminary analysis
• Big data: parallel inference?
• Precise estimation of uncertainty of composite likelihood estimates
• Calibration: how to calibrate test statistics?
• Composite likelihood in hierarchical model?
  ▶ Composite expectation-maximization algorithm
• Can we use composite likelihoods ideas in prediction?
  ▶ Composite kriging?
Thanks!

Merci!

Grazie!
Some references I


Some references II


*Statistics*, **45**.


*Preprint*. University of North Carolina.


Some references III

